TAXONIC

SKOS voor de informatieprofessional

kennis op het web

TAXONIC

Material Subject to Creative Commons License:

A Taxonic whitepaper, by Jan Voskuil @Taxonic taxonic.com

This work is licensed under the Creative Commons Attribution-NonCommercial- NoDerivs 3.0 Unported License. You are free to copy, distribute and transmit the work, under the following conditions: you must attribute the work mentioning the author and Taxonic; you may not use this work for commercial purposes; you may not alter or transform this work.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

For questions regarding usage, please contact info@taxonic.com

TAXONIC

Inhoudsopgave

1 SKOS 4
1.1 Begrippen en semantische relaties 4
1.2 Trefwoorden op het web 4
1.3 Taxonomieën en flexibiliteit 5
1.4 Concrete stappen helpen zetten 6

1 SKOS

Sinds kort staat SKOS (Simple Knowledge Organization System) op de 'pas toe of leg uit'-lijst van Forum Standaardisatie. Daarmee is het een verplichte standaard voor overheidsorganisaties. Wat betekent dat voor de informatieprofessional? SKOS brengt systemen van kennisontsluiting naar het web. Om dat te begrijpen moeten we het vakgebied dat zich bezighoudt met kennisontsluiting, van dichtbij bekijken.

1.1 Begrippen en semantische relaties

Kennisontsluiting begint vaak bij de eigen taal. Elke organisatie hanteert begrippen in een specifieke betekenis. Zo is in het gevangeniswezen de 'badmeester' iemand die nieuwe 'inkomsten' behandelt: inname eigendommen, uitreiken beddengoed, et cetera. De Belastingdienst hanteert weer een heel andere definitie van 'inkomsten'. Kennis van dergelijke organisatie-specifieke begrippen is essentieel. Nieuwe medewerkers dienen zich de terminologie eigen te maken. De definities spelen een cruciale rol in het ontwerpen en bouwen van IT-systemen en zijn een voorwaarde voor het uitwisselen van informatie met andere organisaties. Het is dan ook een goed idee om deze kennis structureel en systematisch vast te leggen en te ontsluiten.

De simpelste manier om begrippen vast te leggen is de alfabetisch gesorteerde woordenlijst. Vanaf de 19de eeuw ontstaat er brede consensus dat een hiërarchische ordening beter werkt. Om de betekenis van 'lente' te doorgronden is het essentieel te weten dat het een nauwer begrip is van 'seizoen', naast 'zomer', 'herfst' en 'winter'. Een woordenlijst die zulke semantische relaties expliciet beschrijft noemen we een thesaurus. Het ontwerp van SKOS laat de betekenis van de relaties doelbewust vaag, waardoor een SKOS-thesaurus een wat informeel karakter krijgt. Daarin schuilt veel van de kracht van deze standaard: de beschrijving is intuïtief, direct toegankelijk voor een breed publiek en biedt tegelijkertijd aantoonbaar meerwaarde.

SKOS maakt het mogelijk om begrippenlijsten die op deze manier gestructureerd zijn (thesauri) via het web toegankelijk te maken in een "Wikipedia-achtige" vorm, zodat elk trefwoord een eigen artikel krijgt met definitie, toelichting en links naar gerelateerde trefwoorden. Zo kun je bij het artikel over 'seizoen' eenvoudig doorklikken naar de 'narrowers' (lente, zomer, herfst, winter).

Eigenlijk zou elke organisatie zijn begrippenapparaat op deze manier moeten publiceren. De fundamentele kennis die daarmee beschikbaar komt, vormt immers de basis voor de informatiehuishouding in een organisatie. Een voorbeeld is het Gegevenswoordenboek migratienketen gegevenswoordenboek-migratieketen, dat elke professional met toegang tot het Rijksweb kan raadplegen.

SKOS ontsluit kennis over begrippen niet alleen via de browser, het zorgt er tegelijkertijd voor dat computerprogramma's de diverse relaties in de thesaurus slim kunnen gebruiken voor allerlei doeleinden. Dat komt doordat SKOS een lid is van de snel groeiende familie van Linked data-standaarden, die het mogelijk maken naast documenten ook gegevens op het web te ontsluiten. Daar zit de crux van SKOS.

1.2 Trefwoorden op het web

Het basisidee van SKOS is dat elk begrip een eigen webadres krijgt. Een fictief voorbeeld is belastingdienst. Het begrip 'inkomsten' zoals het gevangeniswezen dat hanteert krijgt weer een ander webadres. Aan het webadres is meteen te zien welke organisatie verantwoordelijk is voor de definitie van het begrip. Dat is niet alleen simpel en doeltreffend, het geeft ook eenduidigheid op een ongekende schaal. Dat leidt tot allerlei nieuwe toepassingen.

In de praktijk wordt SKOS veel gebruikt voor het publiceren van trefwoordsystemen en taxonomieën. Trefwoordsystemen zijn traditioneel verbonden aan een bibliotheek. Binnen die bibliotheek heeft het systeem betekenis, daarbuiten niet. Met de komst van het web is dat veranderd. Bibliotheken die hun collecties via het web ontsluiten kunnen via datzelfde web hun trefwoordsystemen delen. Dat helpt kennisinstituten hun collecties open te stellen en beter bruikbaar te maken. Ze kunnen trefwoordsystemen van andere bibliotheken op hun eigen content toepassen. Zo kun je met het befaamde, in de medische wereld breed gebruikte MeSH- trefwoordsysteem niet alleen bij de National Library of Medicine terecht, maar bij alle kennisinstituten die MeSH toepassen op hun publicatie-databanken. Gebruikers hoeven daar zelf niets voor te doen: zoekmachines handelen dit automatisch af. Het kunnen ophalen van eenduidig gespecificeerde zoekresultaten bij een ongelimiteerde verzameling bronnen op het web, is een van de kernbeloftes van SKOS.

1.3 Taxonomieën en flexibiliteit

Een andere toepassing van SKOS die sterk opkomt, is het werken met taxonomieën. Een taxonomie is, simpel gezegd, een vaste lijst met begrippen voor het kenmerken van een eenheid van informatie, zoals een document of een gegeven. Daarmee krijgt de informatie toegevoegde waarde: het kenmerk vertelt waar de informatie over gaat, wat voor type informatie het is, tot wanneer het geldig is, et cetera.

De Overheid.nl Web Metadata Standaard (OWMS) laat zien hoe je SKOS op deze manier toepast. Deze standaard, die dateert uit 2010, stelt vast hoe informatie van de Nederlandse overheid op het web gekenmerkt moet worden. Daarmee wordt deze informatie beter vind- en bruikbaar en beter te combineren voor burgers en bedrijfsleven. De standaard definieert negen kerneigenschappen die elke publicatie op een vastgestelde manier moet tonen. Neem bijvoorbeeld de eigenschap ‘informatietype’. De waarde van deze eigenschap geeft aan om wat voor soort informatie het gaat. Daaraan kun je het type overheidsdocument herkennen, zoals een omgevingsvergunning, een structuurplan of een beschikking. De makers van OWMS hebben een specifieke SKOS-thesaurus gemaakt die de bij de Nederlandse overheid voorkomende informatietypes inventariseert.

Het gebruik van SKOS zorgt er niet alleen voor dat de identiteit van elke begrip in de thesaurus ondubbelzinnig vaststaat en de betekenis ervan eenvoudig is op te zoeken. Het zorgt ook voor een hoge mate van flexibiliteit. Ten eerste kan de thesaurus uitgebreid worden zonder dat dat nadelige effecten hoeft te hebben op IT-systemen. Bovendien voorziet SKOS in middelen om begrippen in de ene thesaurus te linken aan begrippen in een andere thesaurus. Dat gaat volgens een standaardrecept. Voor organisaties die met elkaar samenwerken is het handig om elkaars definities altijd bij de hand te hebben, ook over landsgrenzen heen. De Europese Unie (EU) heeft een aantal belangrijke thesauri laten maken met begrippen in 23 EU-talen. Overheidsorganisaties kunnen de begrippen in hun eigen thesaurus eenvoudig linken aan begrippen in deze EU-thesauri. Dat bevordert het delen van informatie op Europees niveau.

Een derde instrument dat SKOS aanreikt om OWMS flexibel te houden is dat partijen in een specifiek domein ervoor kunnen kiezen om een andere, gespecialiseerde thesaurus te gebruiken voor de waarden van een eigenschap als ‘informatietype’. Doordat de waarde de vorm heeft van een webadres dat naar het begrip verwijst, worden misverstanden voorkomen.

Het op deze manier kenmerken van digitale documenten biedt grote voordelen. Documenten uit het ene IT-systeem worden bruikbaar in het andere IT-systeem. Stromen van documenttypen over organisaties heen wordt daardoor inzichtelijker en transparanter en leveren minder ICT-kosten op

voor de dienstverlener. De afnemer kan zijn digitale papierwinkel eenvoudig ordenen, langs diverse assen. Een overzicht van alle vergunningen van een bepaald type is dan een kwestie van een druk op de knop. Voor een aannemer in glasvezelprojecten die duizenden vergunningen per jaar binnenkrijgt

zou dat een grote stap vooruit betekenen. Het automatisch kunnen verwerken van omvangrijke digitale documentstromen levert namelijk flinke besparingen op.

1.4 Concrete stappen helpen zetten

Tegen de tijd dat SKOS breed geadopteerd is, zijn thesauri via het web verbonden met andere thesauri. Begrippenkaders zoals die binnen een organisatie gebruikt worden zijn eenvoudig met die van andere organisaties te vergelijken, ook internationaal. Thesauri worden gebruikt en hergebruikt voor kennisdeling, trefwoordtoekenning, metadatering, automatische tekstanalyse, en het verbinden van informatie. Administratieve processen lopen soepeler, de resultaten zijn eenduidiger en transparanter. Publicatiesystemen en zoekmachines ondersteunen SKOS en kunnen met elke willekeurige thesaurus overweg. IT-Systemen zowel als menselijke gebruikers gebruiken ze simpelweg via een hyperlink. Begrippen en definities zijn toegankelijk zonder drempels, informatie uit verschillende bronnen is beter vindbaar en vergelijkbaar.

Opname van SKOS op de ‘pas toe of leg uit’-lijst van Forum Standaardisatie brengt deze toekomstdroom een stapje dichterbij. De meer dan twintig partijen die zich actief hebben ingezet om dit mogelijk te maken hopen dat het verplicht stellen van SKOS organisaties stimuleert om hun begrippen vast te leggen en conform SKOS te ontsluiten. Een drempel die daarbij genomen moet worden is dat de echte baten daarvan pas blijken als veel organisaties dat doen. Het is dus verleidelijk om in de wachtstand te blijven hangen. Ook leidt SKOS tot meer transparantie, wat organisaties vaak onzeker maakt. Het valt te hopen dat een stok achter de deur de balans een duwtje in de goede richting geeft.

Taxonic

Wij zijn Taxonic, een internationaal en onafhankelijk IT-adviesbureau. Onze focus? Digitale transformatie en Kennismanagement, dit passen wij toe binnen twee gebieden: Pega (Dynamic Case Management) en Linked Data.

Als kennispartner delen we onze kennis, geven we advies op maat en zijn we hands-on op projecten van analyse tot en met implementatie. Onze experts werken met de nieuwste technologieën. Binnen het team hebben we een brede ervaring; van advies over de keuze van technologie en het inbedden van oplossingen in de technische omgeving tot de koppeling van het systeemlandschap en volledige realisatietrajecten.

Doordat wij werken met interdisciplinaire teams kunnen wij een breed spectrum aan complexe projecten oppakken. Taxonic stimuleert de wereld om mee te veranderen met nieuwe technologieën, dat is onze bijdrage aan de maatschappij.

Jan Voskuil

Met een achtergrond als taalwetenschapper begrijpt Jan de begrippenstructuur en het belang van Linked Data en SKOS als geen ander. Hij heeft zich daarom, vanuit een inhoudelijke drijfveer, hard gemaakt voor de ontwikkeling van Linked Data in Nederland. Jan leest graag en is geïnteresseerd in geologie; dat is geen vreemde combinatie voor een informatie architect.

Jan is een “thought leader” met diverse publicaties op zijn naam over Semantische technologie en Linked Data. Ook heeft hij SKOS in Nederland op de kaart gezet door deze zo te agenderen dat hij wordt opgenomen in de ‘pas-toe-of-leg-uit’-lijst.

The information in this document is the property of Taxonic BV. Copyright ©2023 Taxonic. All rights reserved.

Veldzigt 2
3454 PW De Meern
The Netherlands

T +31 (0)88 829 66 00
E info@taxonic.com
W www.taxonic.com

bank NL45RABO0161959660
BTW NL851339803B01
KvK 54529190